Moving Large Data to Galaxy

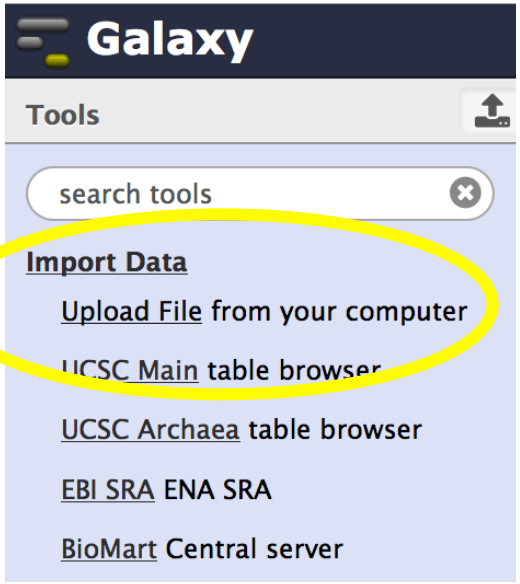
Tom Doak
Le-Shin Wu
Carrie Ganote
National Center for Genome Analysis Support

July 16, 2014

INDIANA UNIVERSITY

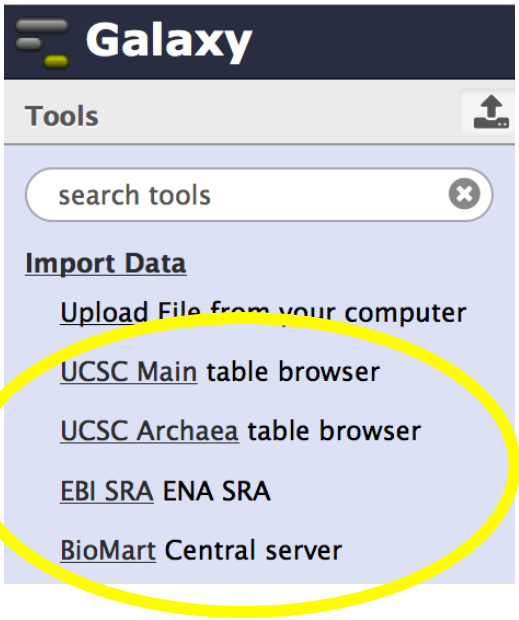
Let's get some sequence data

Moving large data sets onto Galaxy can be done in a number of ways:



For small files, use the Import Data -> Upload File from your computer. The page states a limit of 2GB, but in practice I wouldn't upload a file larger than a few MB this way - It's slow and tends to silently crash.

Let's get some sequence data



Data sources from UCSC and other browsers allow you to pull specific data from public resources

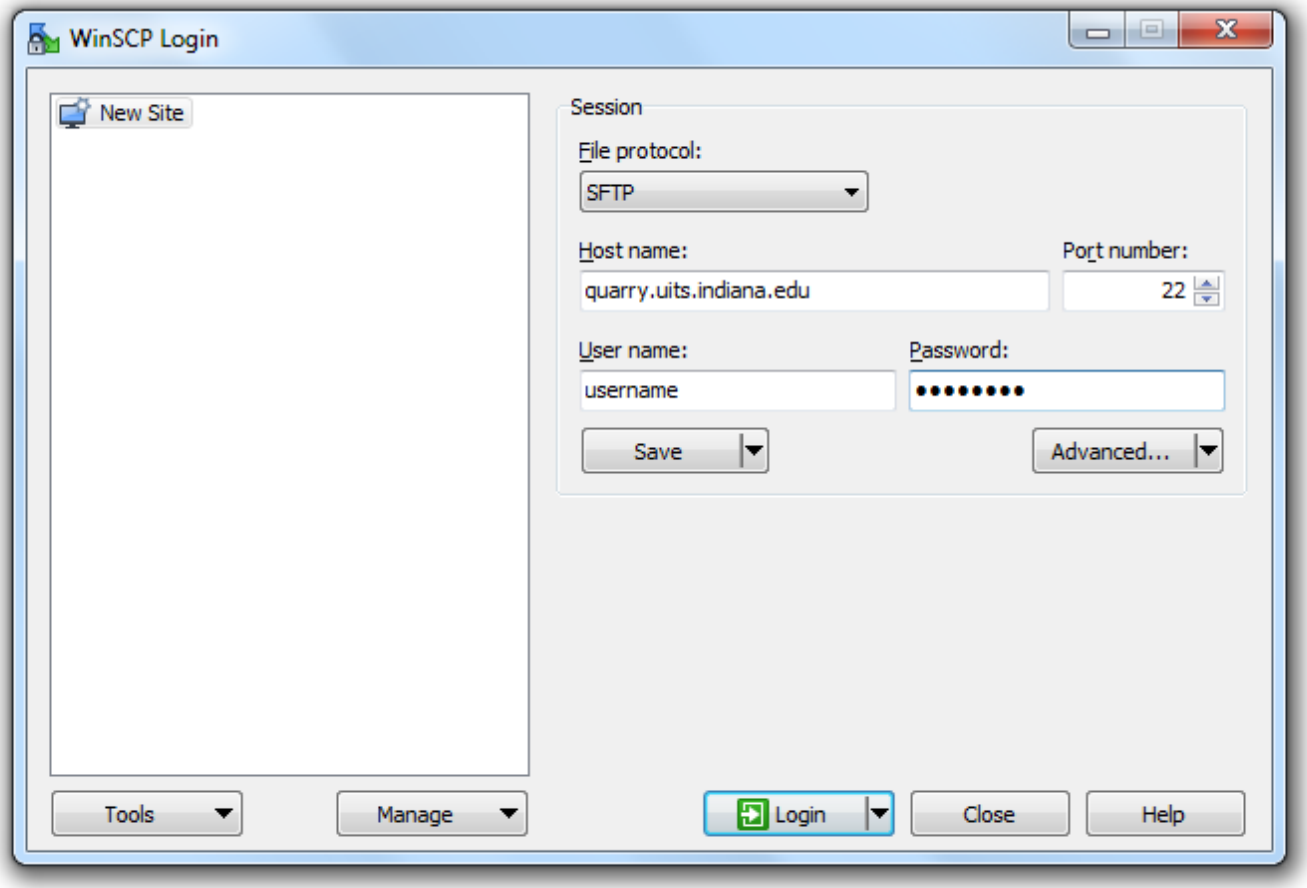
Let's get some sequence data

The best solution for large files is to move files directly to the server

Any SFTP, SCP, or Globus client can be used to move files – we recommend WinSCP and Cyberduck

You will need an account on Mason, Quarry, or Big Red2. You can request accounts at itaccounts.iu.edu.

Let's get some sequence data



WinSCP

Login to host:

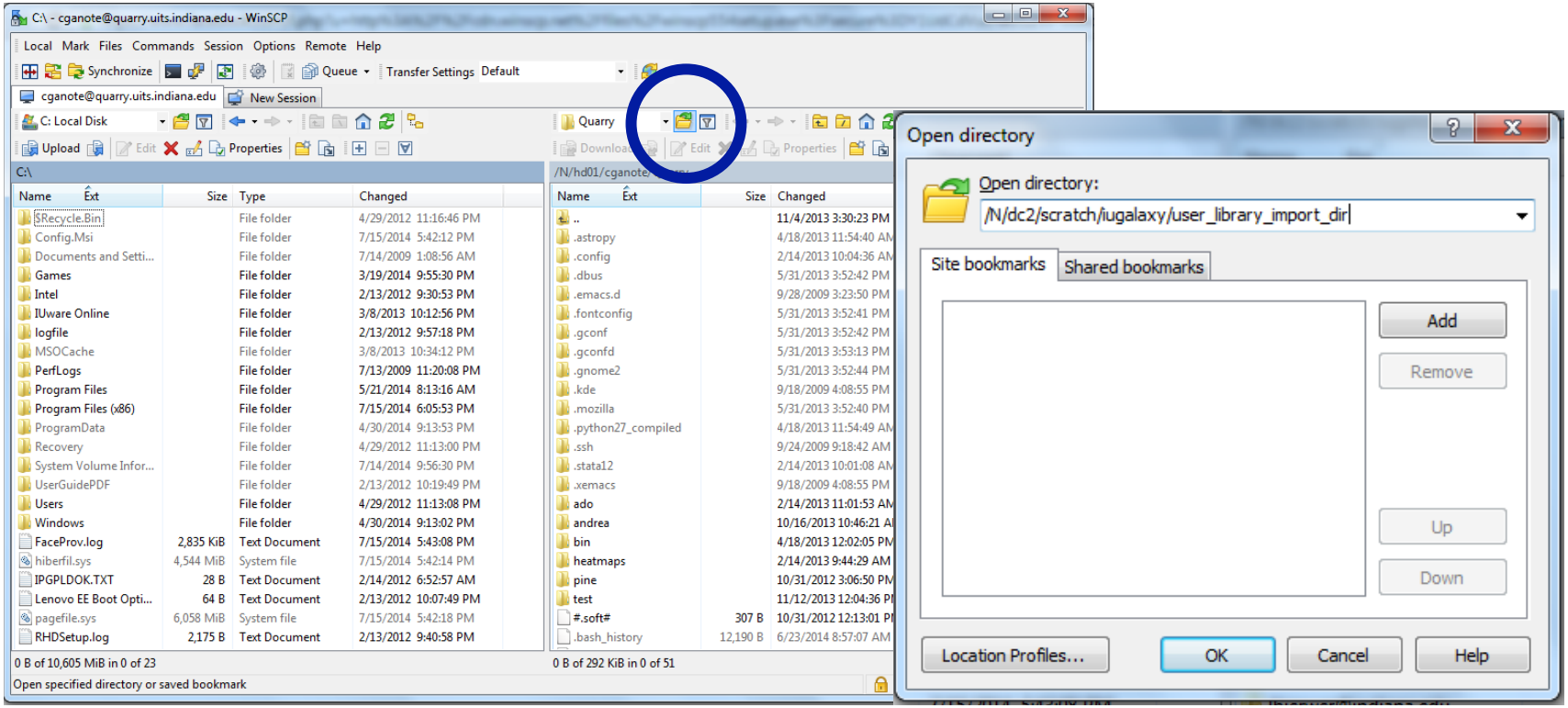
quarry.uits.indiana.edu

mason.indiana.edu

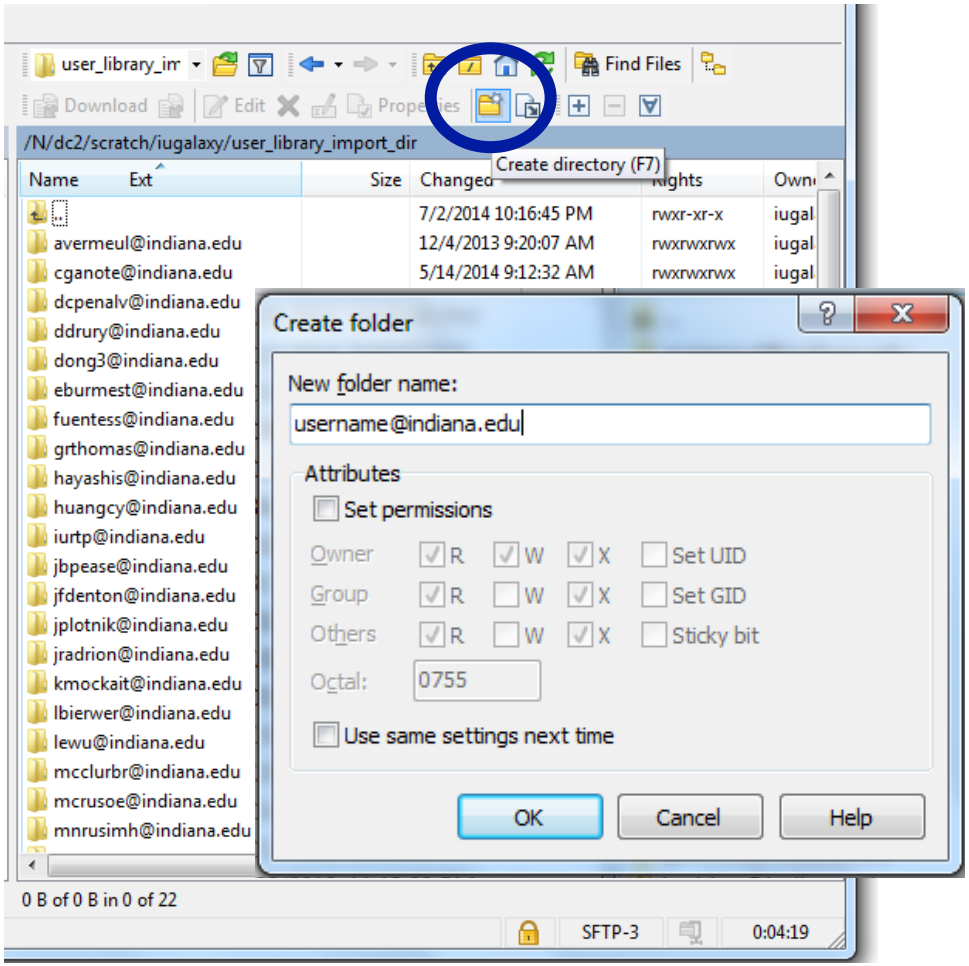
bigred2.iu.edu

Let's get some sequence data

Open directory: /N/dc2/scratch/iugalaxy/user_library_import_dir

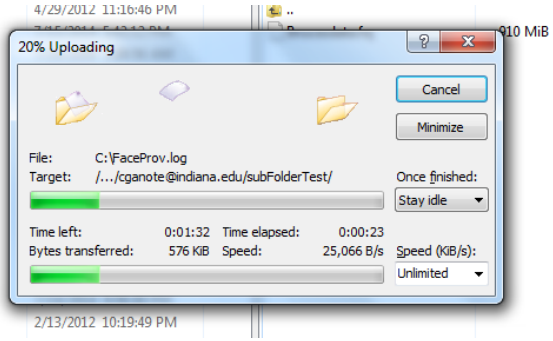


Let's get some sequence data

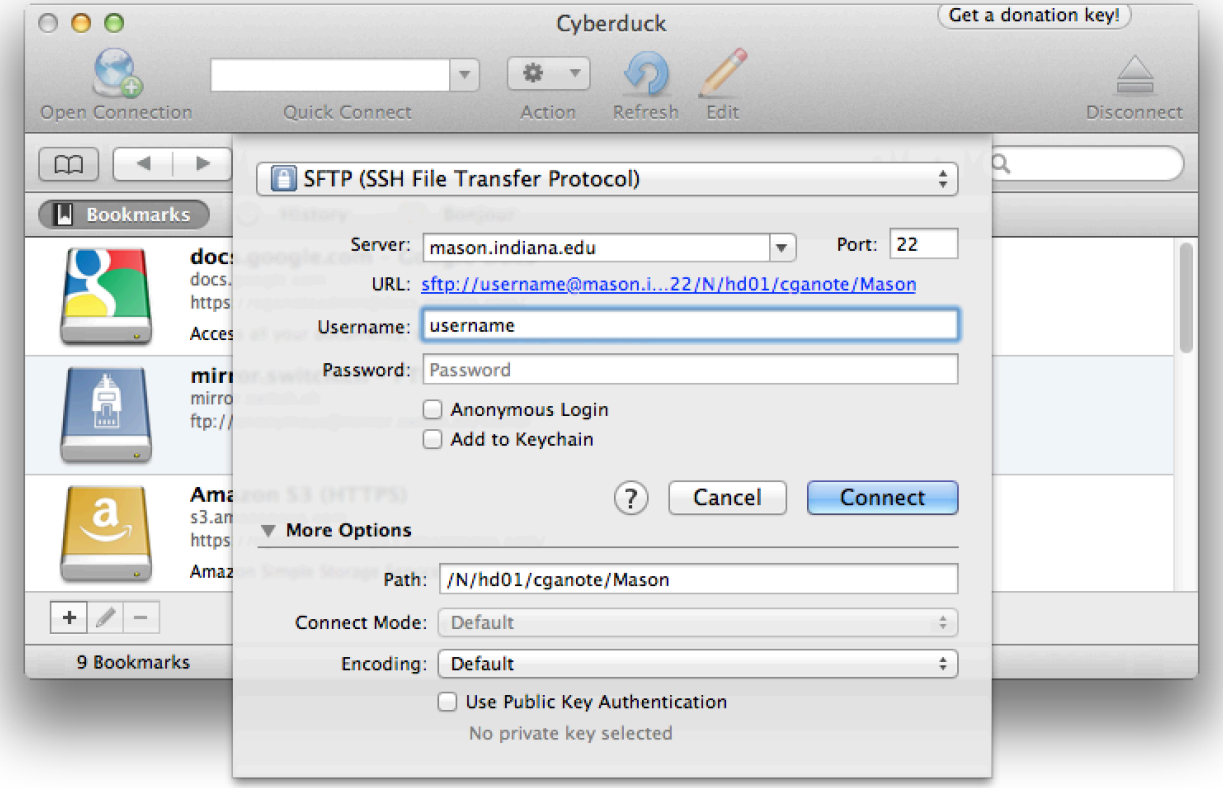


Create a folder if you don't have one: your username@indiana.edu

Then it is a matter of dragging and dropping files!



Let's get some sequence data



Cyberduck

Login to host:

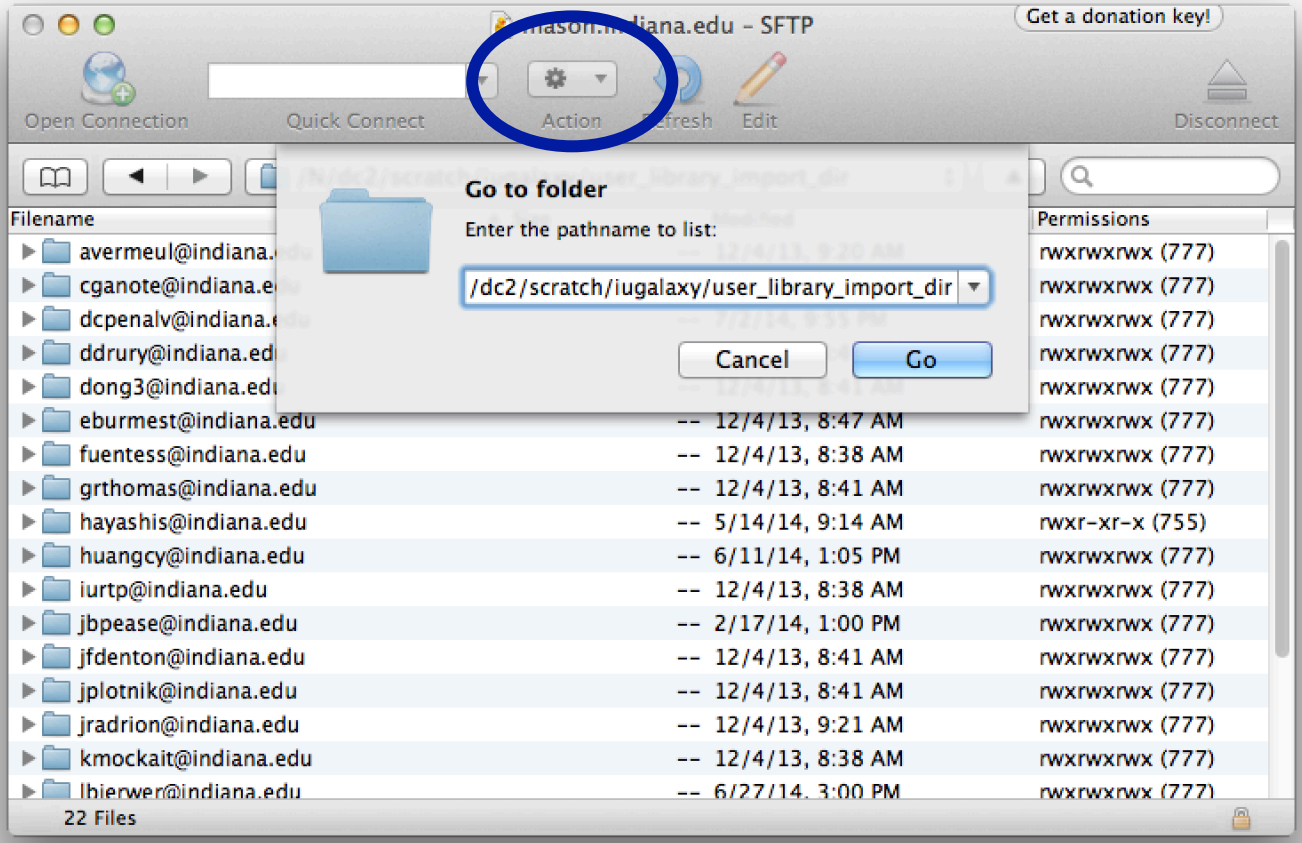
quarry.uits.indiana.edu

mason.indiana.edu

bigred2.iu.edu

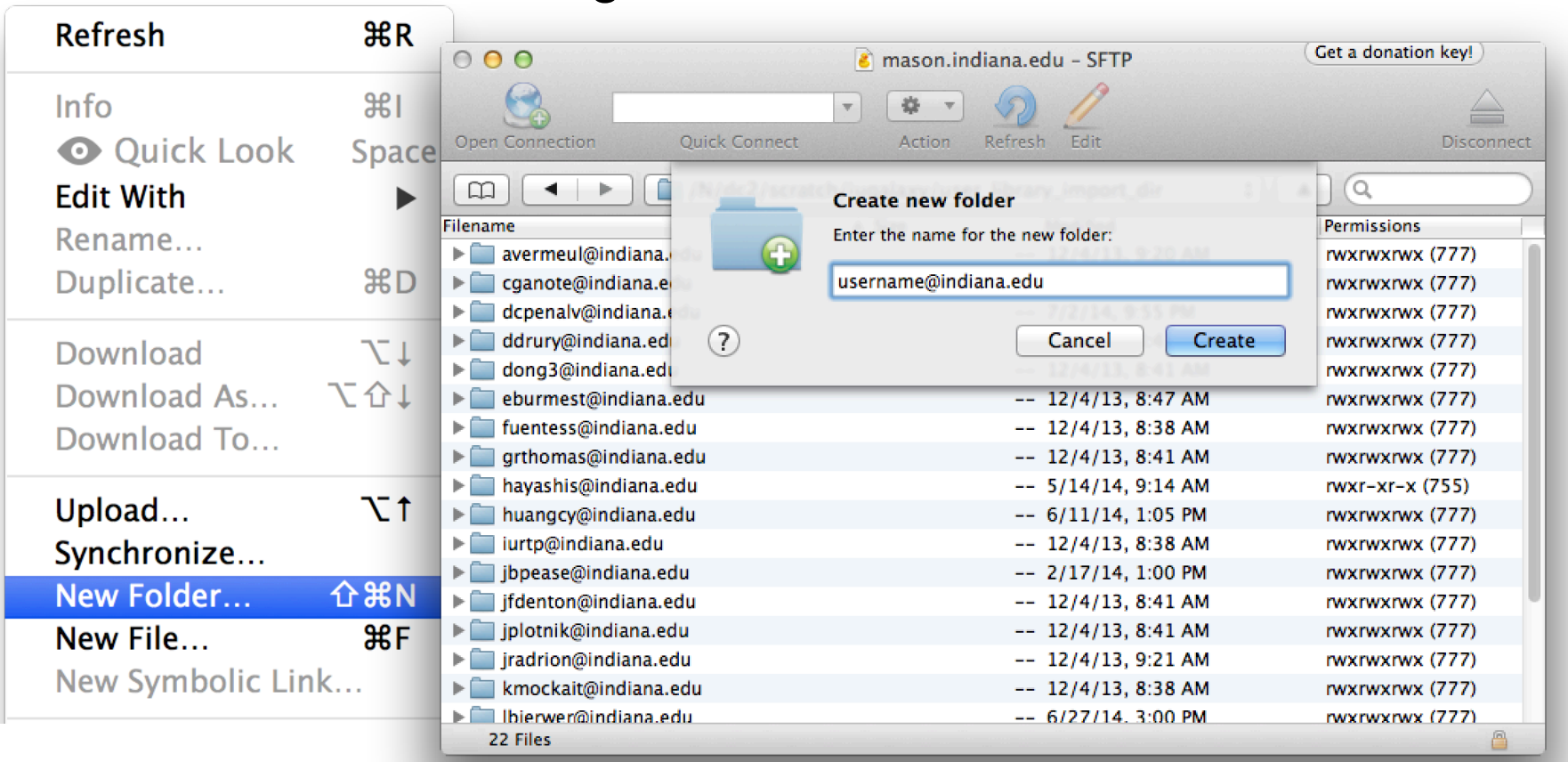
Let's get some sequence data

Actions:
Go to folder:
Same as before

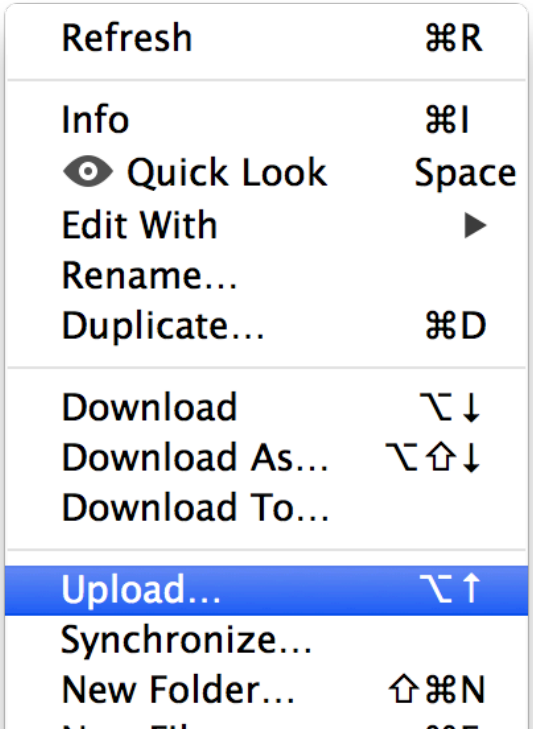


Let's get some sequence data

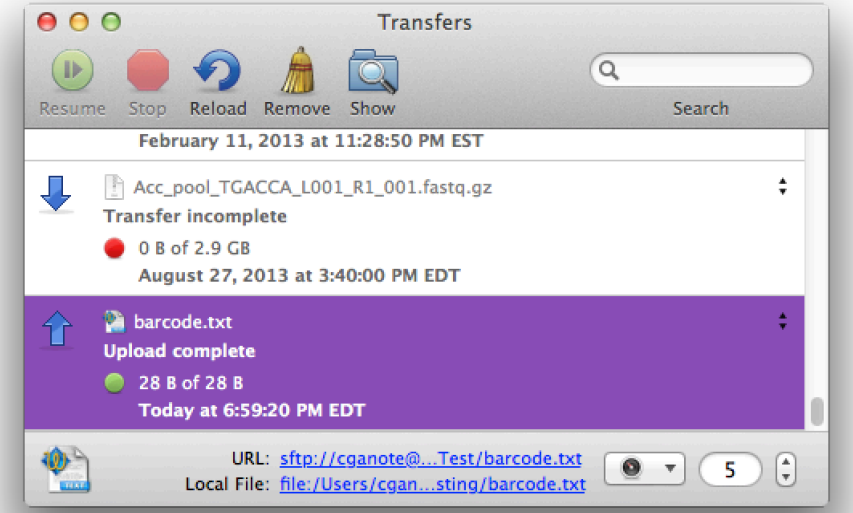
Right click to create a new folder



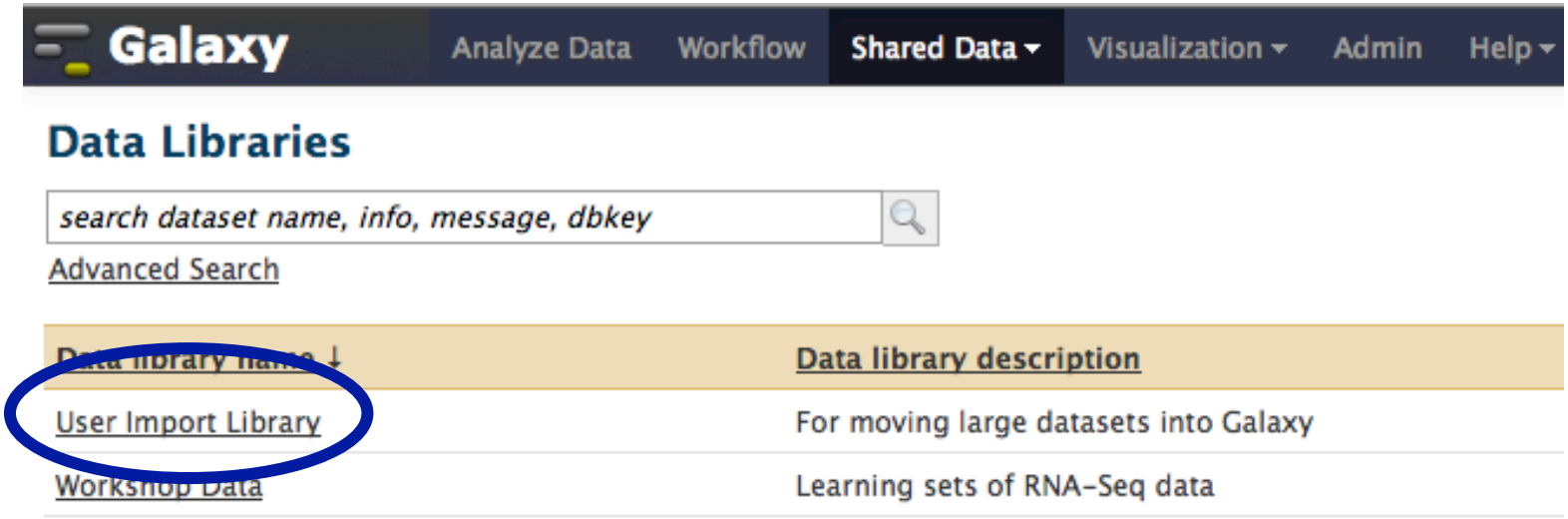
Let's get some sequence data



Cyberduck only shows the remote view, so when you are in the folder you want, right click and choose Upload to browse local files

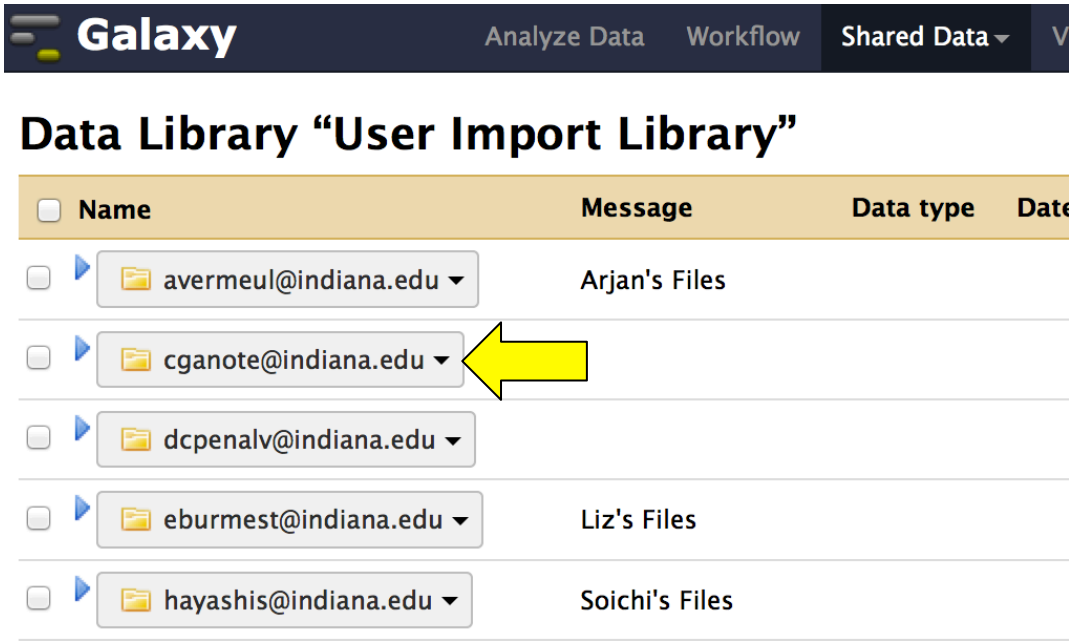


Let's get some sequence data



Go to Shared Data -> Data Libraries -> User Import Library

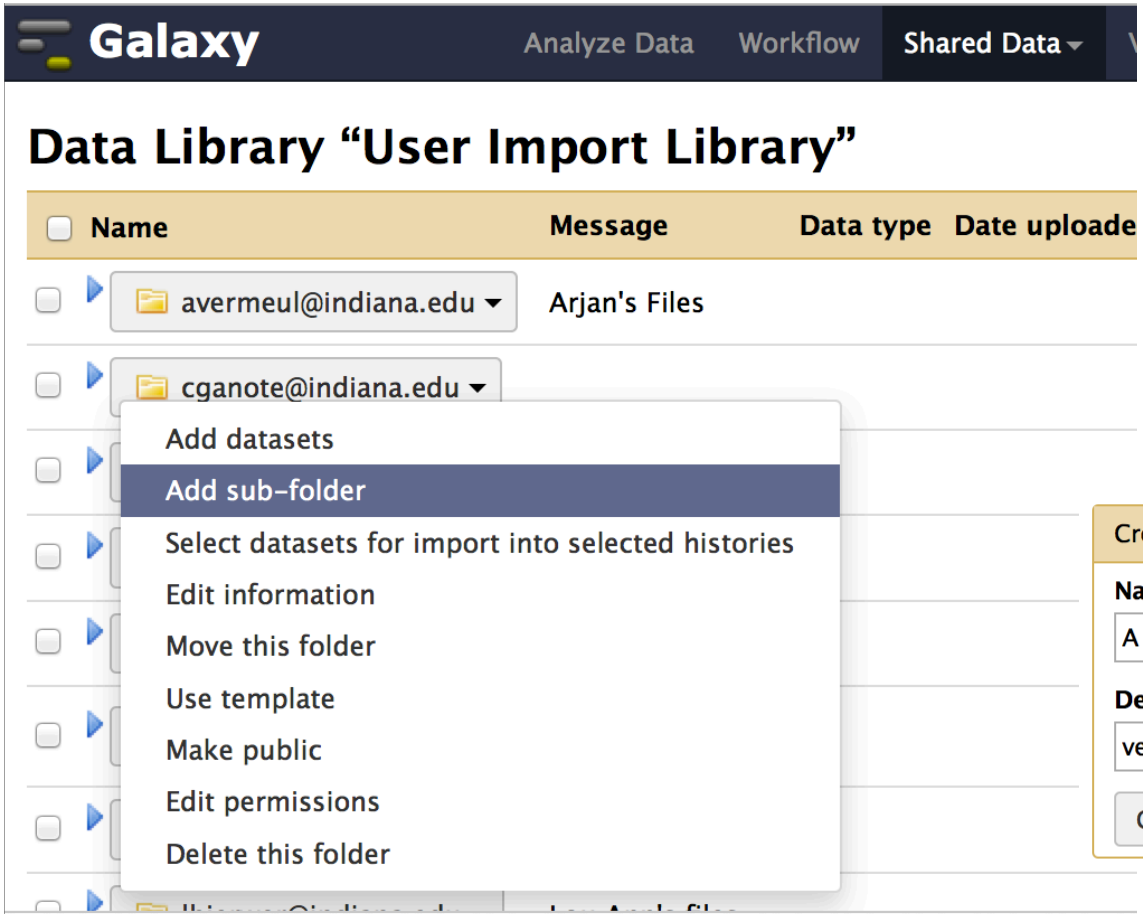
Let's get some sequence data



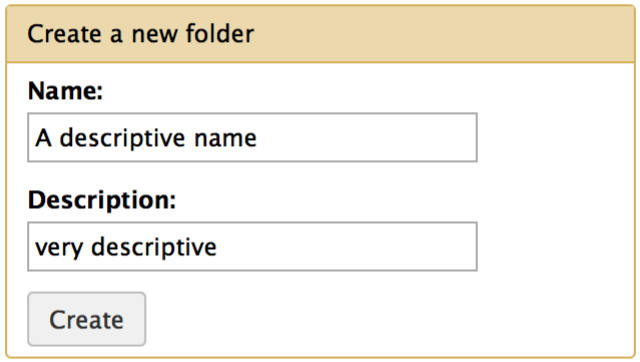
Find your username – click on the black arrow to right of it

If you don't see your name, contact us at help@ncgas.org and we will add you ASAP!

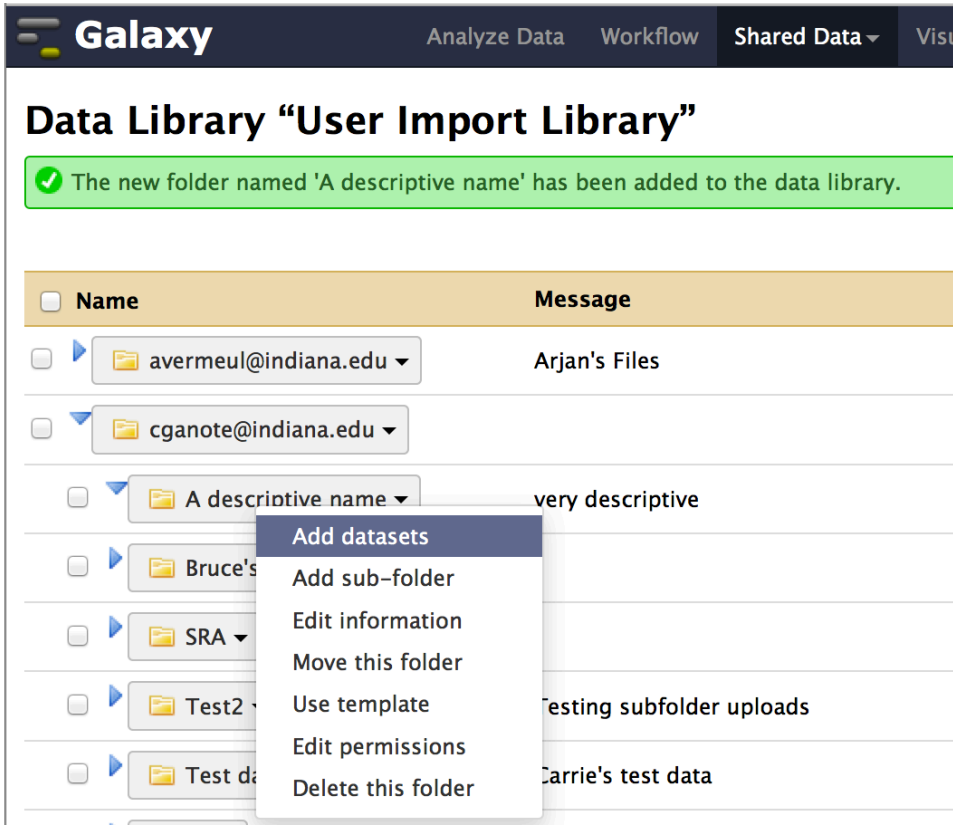
Let's get some sequence data



Add a sub-folder with a descriptive name



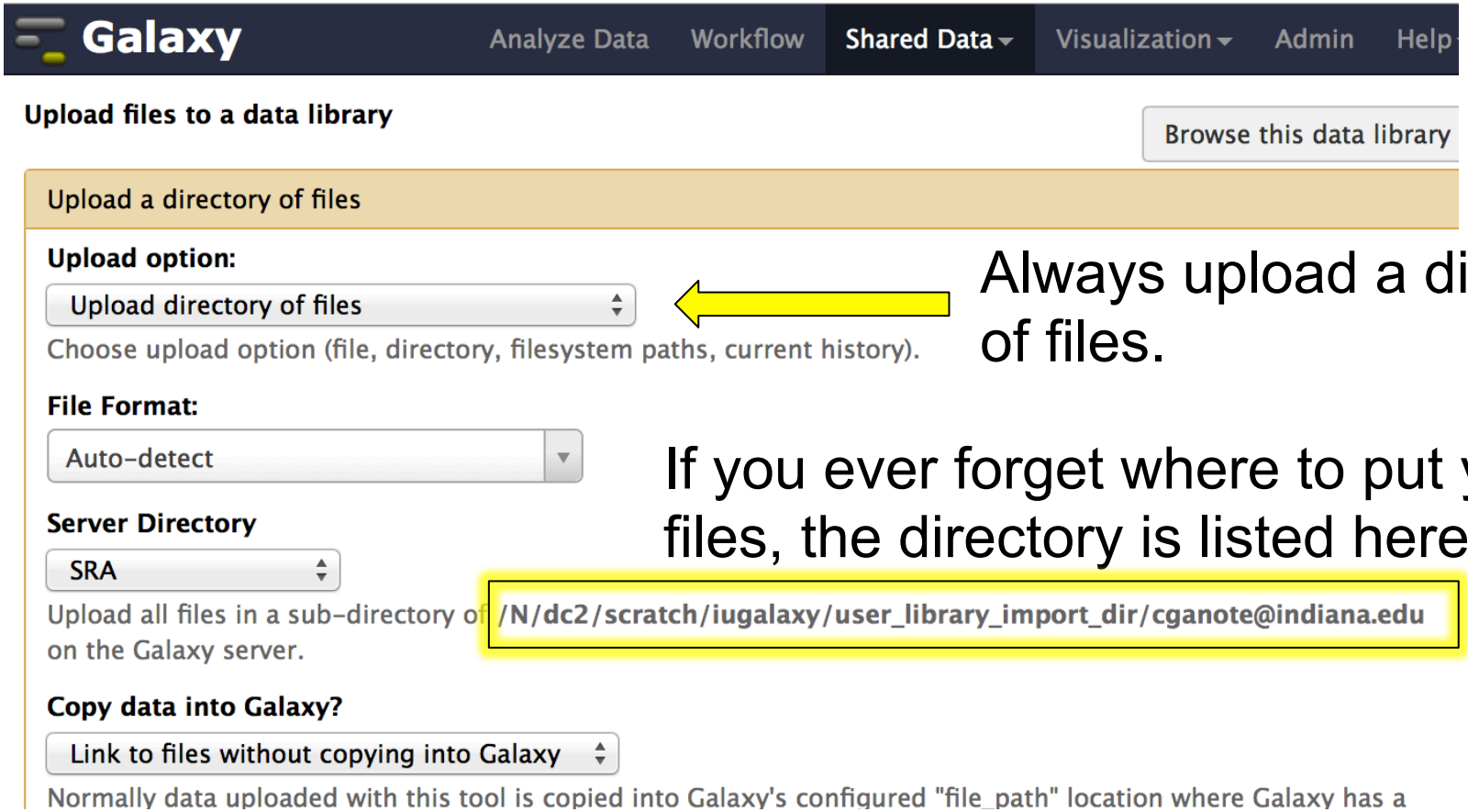
Let's get some sequence data



Right click on the black arrow to the right of the new folder.

Choose to Add datasets.

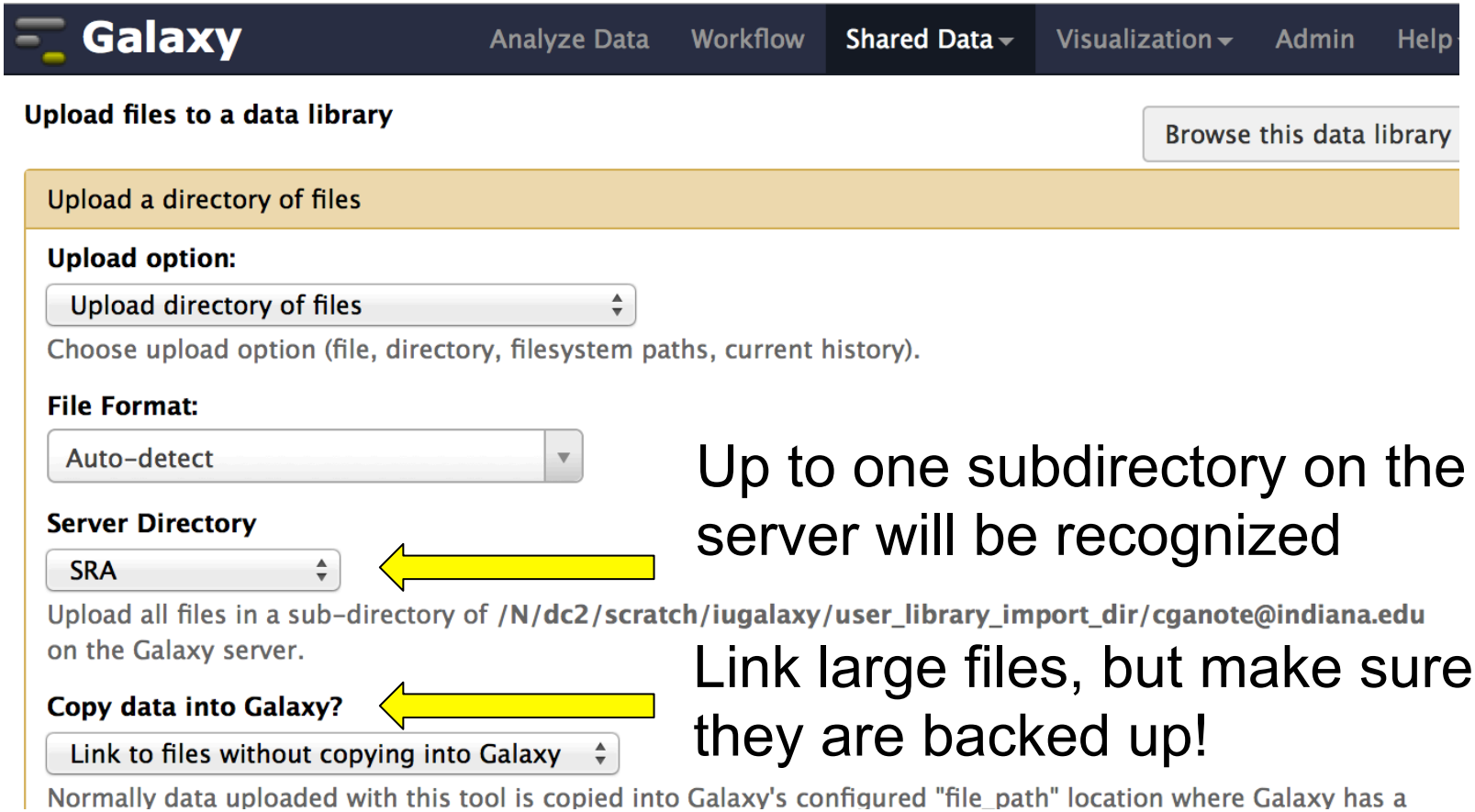
Let's get some sequence data



Always upload a directory of files.

If you ever forget where to put your files, the directory is listed here:

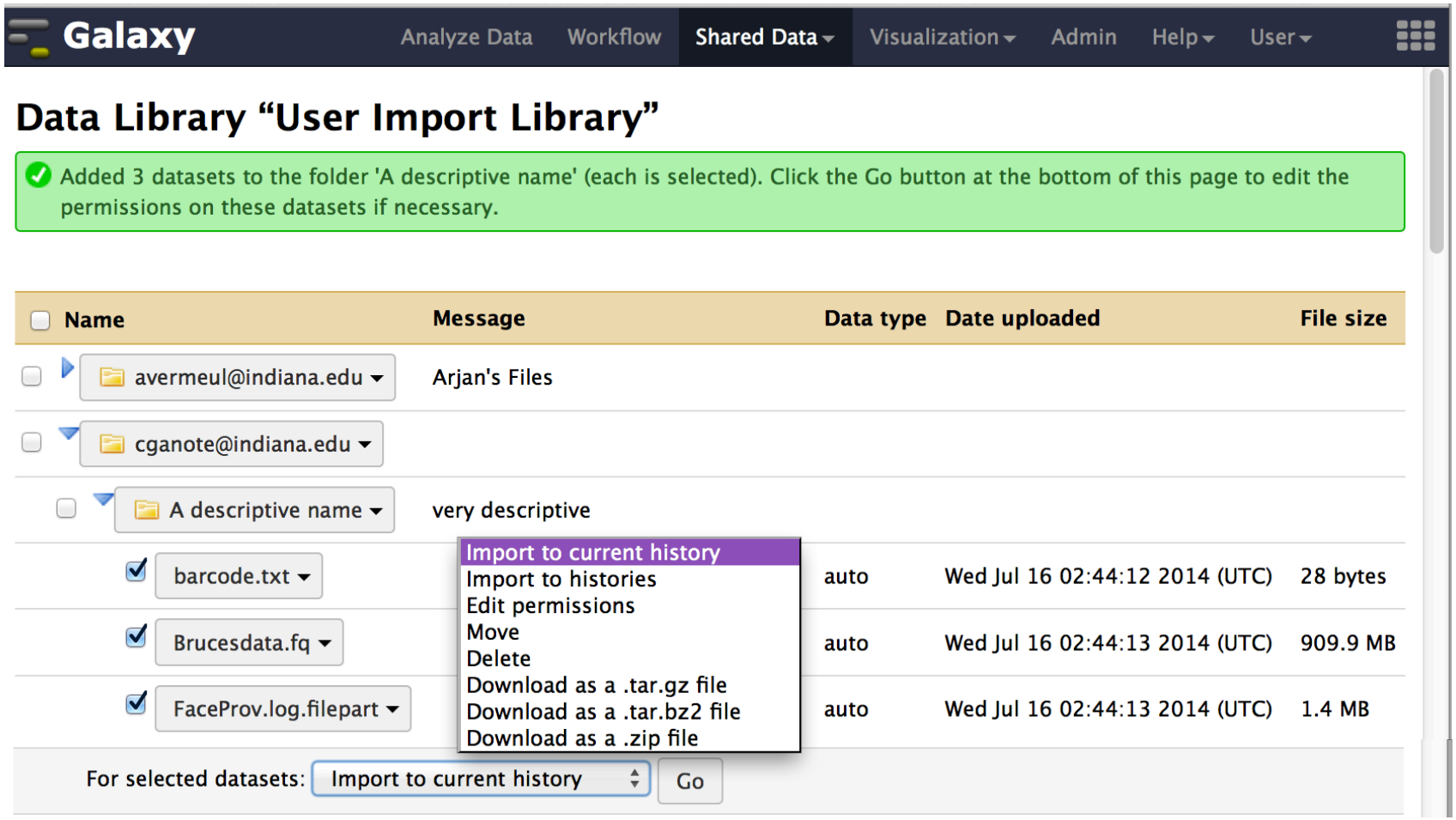
Let's get some sequence data



Up to one subdirectory on the server will be recognized

Link large files, but make sure they are backed up!

Let's get some sequence data



Fín

Thanks for watching!

Questions and comments:

Email help@ncgas.org